

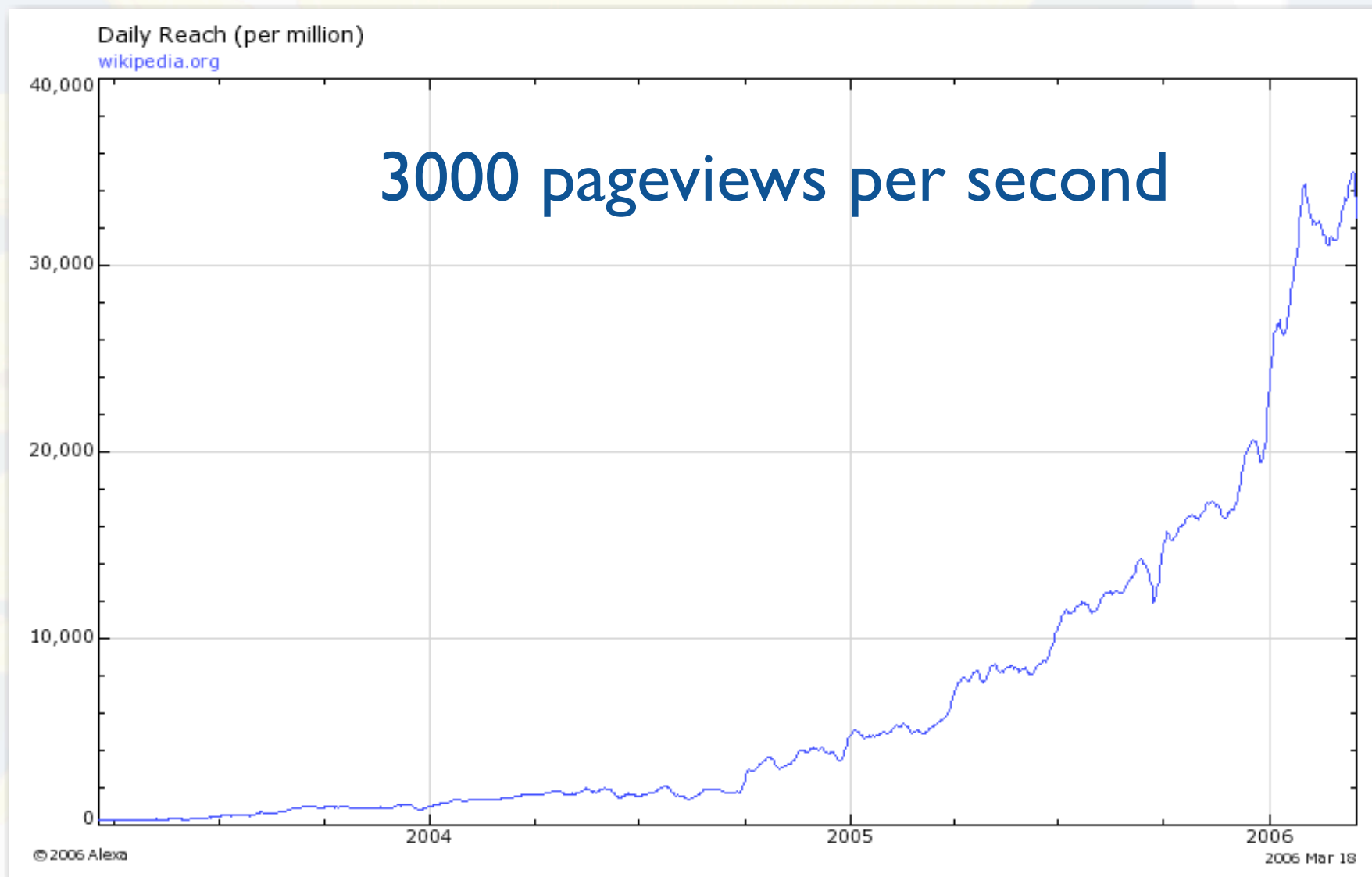
Wikipedia and MediaWiki

What are those crazy wiki people up to anyway?

Servant of two masters

- To the public:
Easy, freely accessible reference work...
always up to date!
- To the Wikipedia community:
MMORPG text adventure...
pages always changing!

To date, most tech effort spent on scaling



To date, most tech effort spent on scaling

- Database replication, multiple-read single-write
- Explicit checks for slave synchronization lag
- Shared cluster caches in memcached
- Explicit control of Squid reverse proxy caches
- *and more of that sort of boring stuff*

Where do we go from here?

- **Share the love!**
You know, all that Web 2.0 crap
- **Make it work!**
Friendlier failure modes
- **Make it fast!**
Avoid click-n-wait

- Continuous integration: code is always ready to run Wikipedia
- Switched CVS -> Subversion
Easier branch management for work on big changes
- Quarterly releases of MediaWiki:
 - 1.6 - April 2006
 - 1.7 - July 2006
 - 1.8? - October 2006
 - ...

Time to work on quality user experience

What's it all about?



“So how’s a *wiki*
different from a *blog*?”

Googlefight

The classics

Funny fights

Fight of the month

The last 20 fights



Results on Google :

Blog

1,240,000,000 results

Wiki

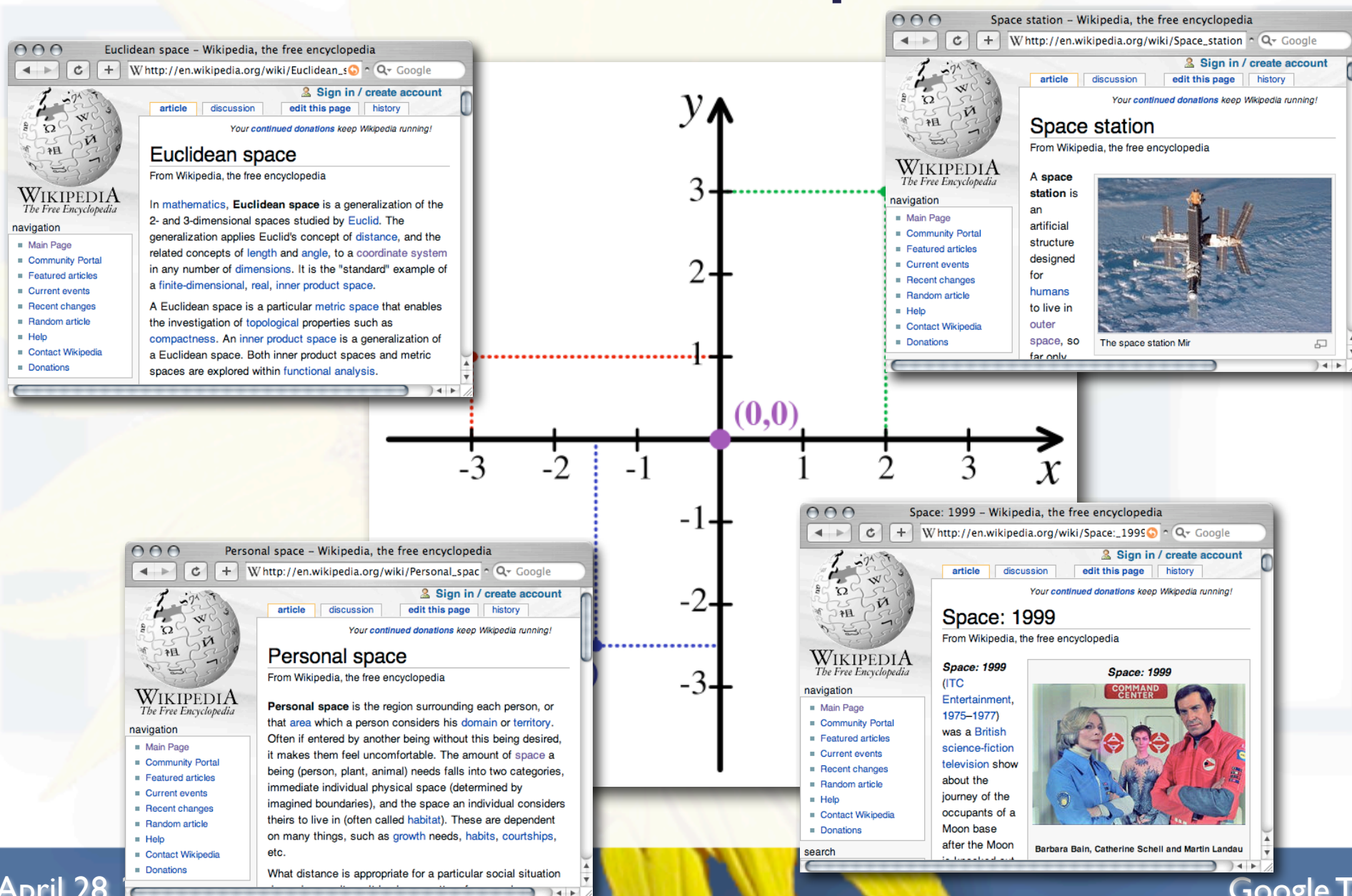
507,000,000 results

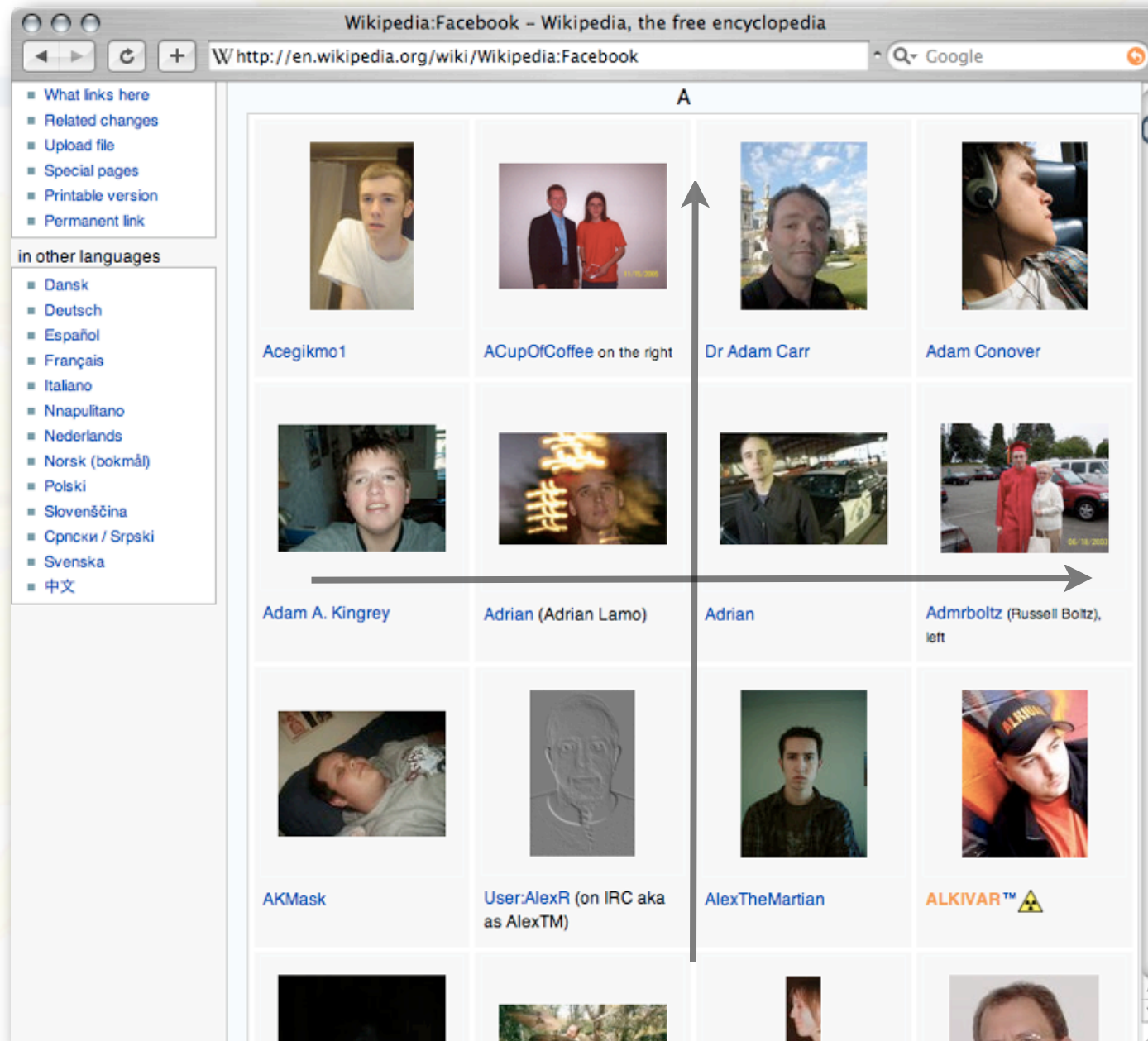


Blogs exist in Time

2006	Friday, April 28	
8 AM		
9 AM		
10 AM	Pack up conference materials, check out of hotel	
11 AM		
Noon		
1 PM		
2 PM	Give talk at Google	
3 PM		
4 PM		
5 PM		
6 PM	Mope on LiveJournal or whatever it is bloggers do	

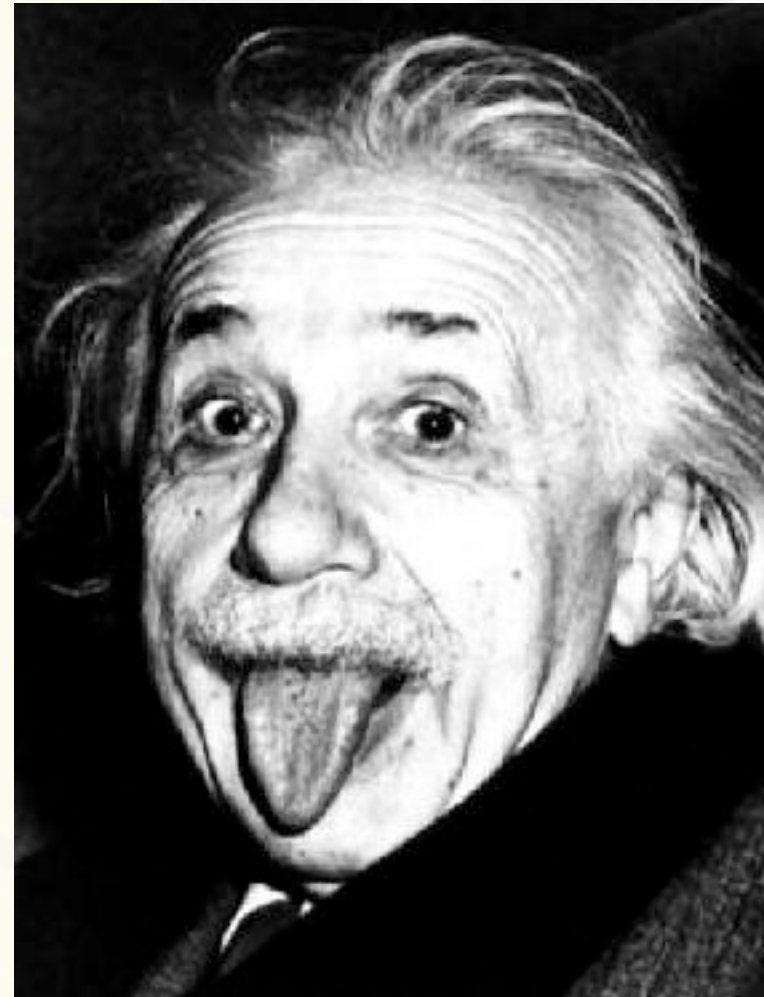
Wikis in space



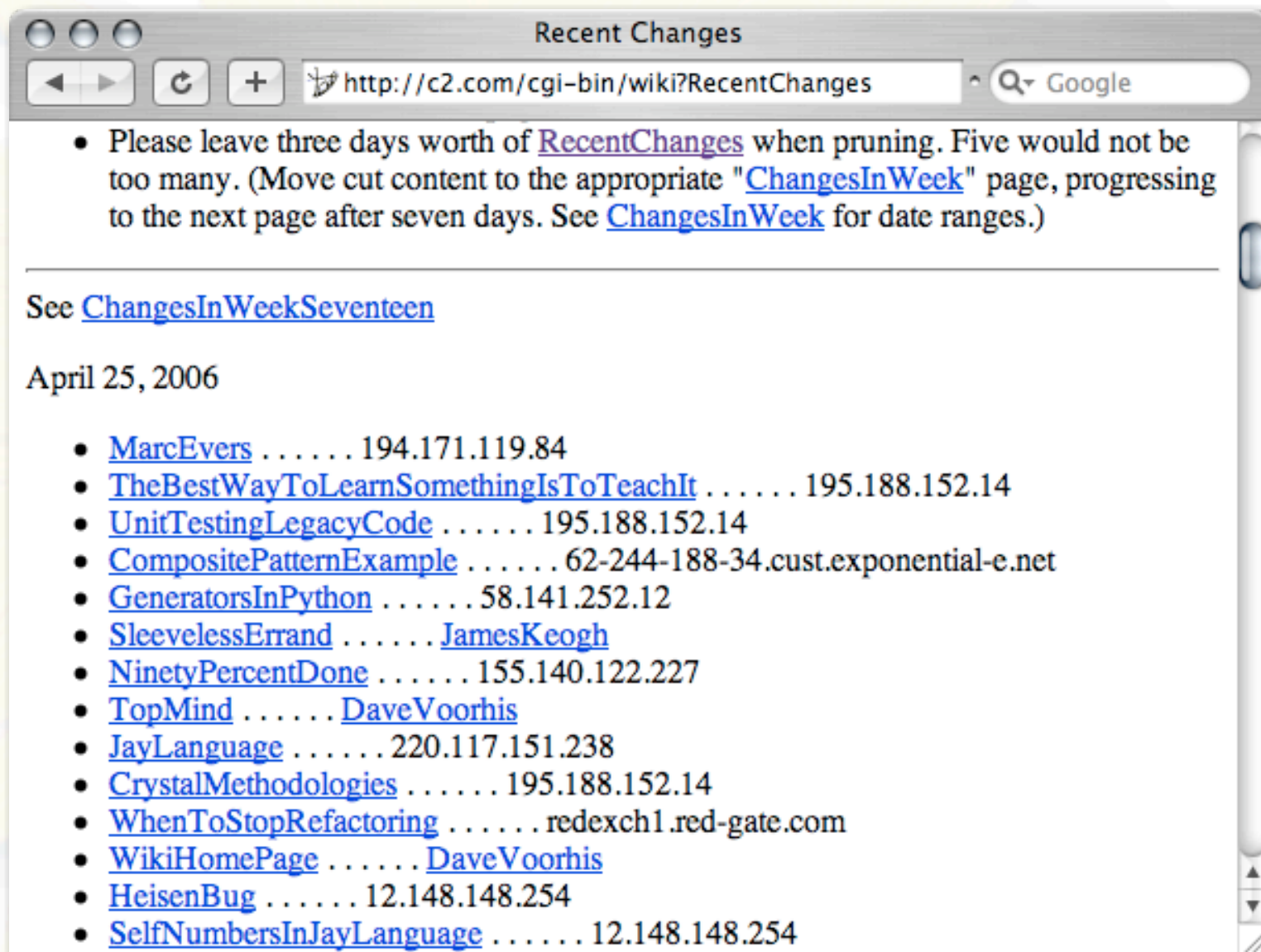


also in
people-
space

What about Spacetime?



Wikis have time too



Lots of time



George W. Bush - History - Wikipedia, the free encyclopedia

W http://en.wikipedia.org/w/index.php?title=George_W._Bush&action=history Google

[Sign in / create account](#)

[article](#) [discussion](#) [view source](#) [history](#)

Your *continued donations* keep Wikipedia running!

George W. Bush

From Wikipedia, the free encyclopedia

[Revision history](#)

(Latest | [Earliest](#)) View (previous 1000) ([next 1000](#)) ([20](#) | [50](#) | [100](#) | [250](#) | [500](#)).

To view a previous version, click the date for that version.

Legend: (cur) = difference with current version, (last) = difference with preceding version, **m** = minor edit

[Compare selected versions](#)

- (cur) (last) ☒ 08:49, 28 April 2006 [Chrisincanadaoraustralia](#) **m**
- (cur) (last) ☒ 07:39, 28 April 2006 [Wechselstrom](#) **m** (*→ Supreme Court nominations / appointments - corrected grammar and clarity*)
- (cur) (last) ☐ 06:53, 28 April 2006 [Voice of All](#) **m** (*Reverted edits by [Peterfa](#) ([talk](#)) to last version by [DHN-bot](#)*)
- (cur) (last) ☐ 06:29, 28 April 2006 [Peterfa](#)
- (cur) (last) ☐ 04:27, 28 April 2006 [DHN-bot](#) **m** (*robot Modifying: ru*)
- (cur) (last) ☐ 03:44, 28 April 2006 [Weirdkid](#) **m** (*It isn't "supposed" if it is self-admitted and documented as having occurred.*)
- (cur) (last) ☐ 02:28, 28 April 2006 [That Guy, From That Show!](#) **m** (*Reverted edits by [SuperDeng](#) ([talk](#)) to last version by [Jaranda](#)*)
- (cur) (last) ☐ 02:27, 28 April 2006 [SuperDeng](#) (*→ See also - restored information that was lost during the restoration*)

WIKIPEDIA
The Free Encyclopedia

navigation

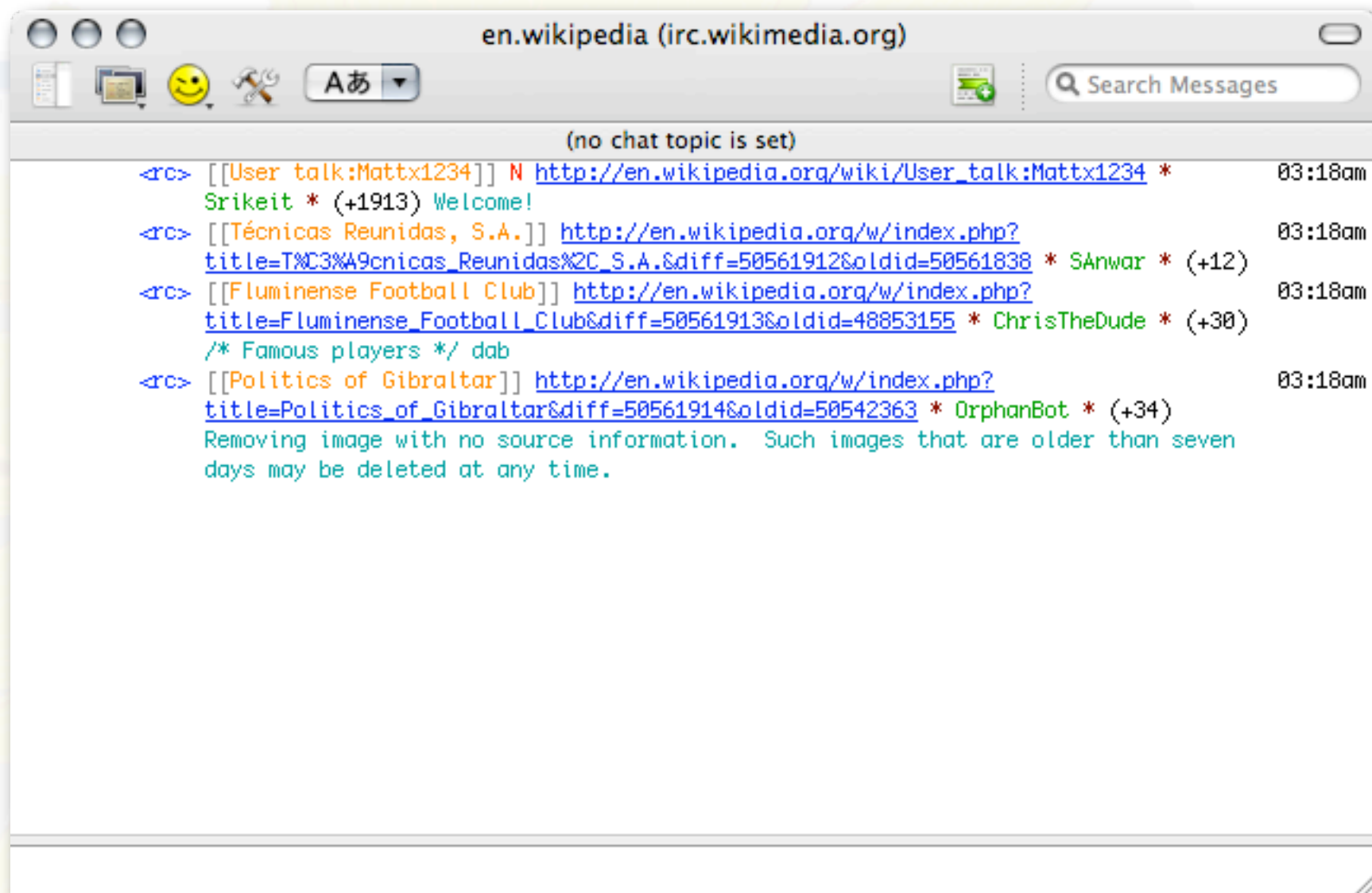
- [Main Page](#)
- [Community Portal](#)
- [Featured articles](#)
- [Current events](#)
- [Recent changes](#)
- [Random article](#)
- [Help](#)
- [Contact Wikipedia](#)
- [Donations](#)

search

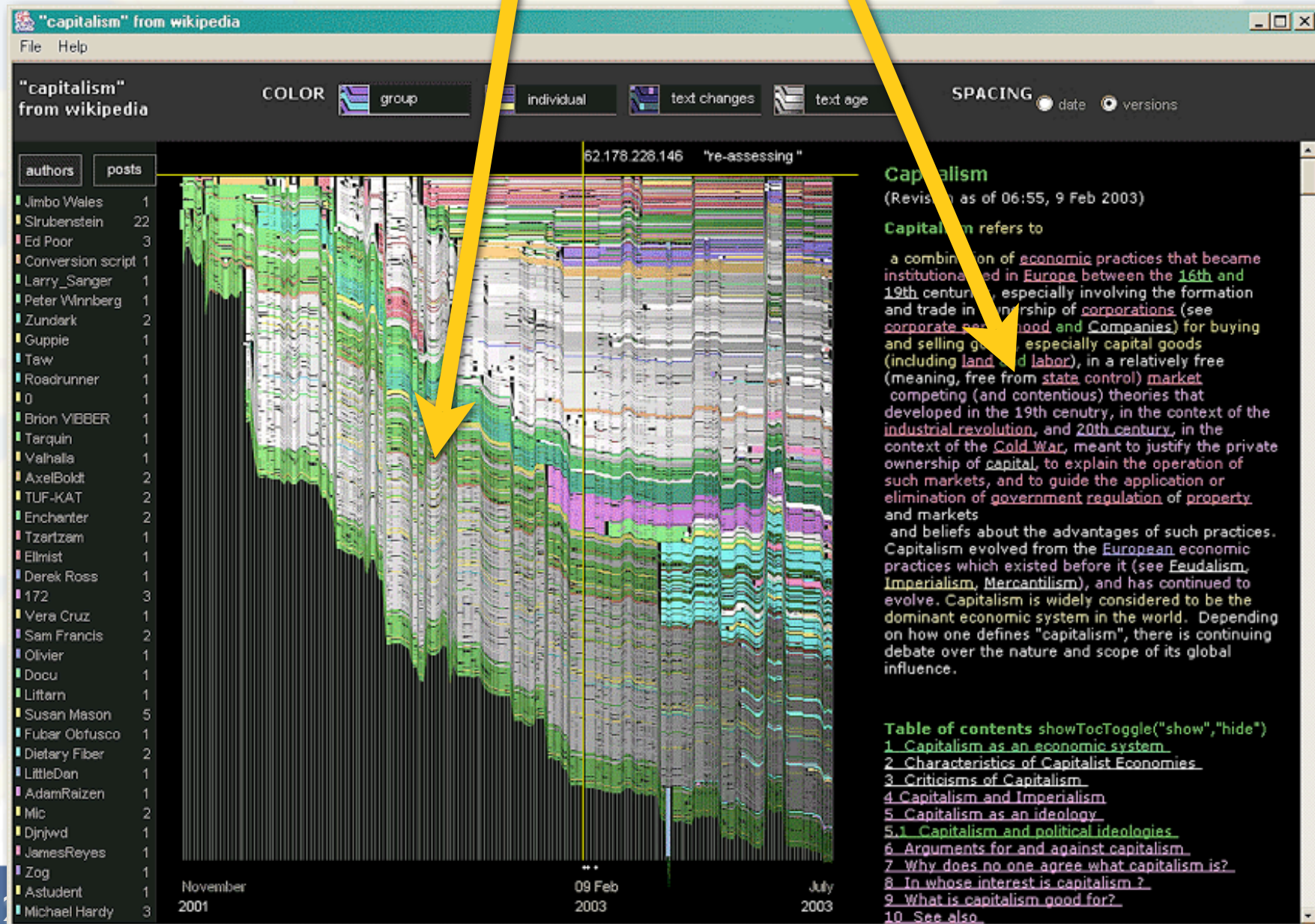
[Go](#) [Search](#)

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)



time = space





Compressing time, compressing space

search filters space



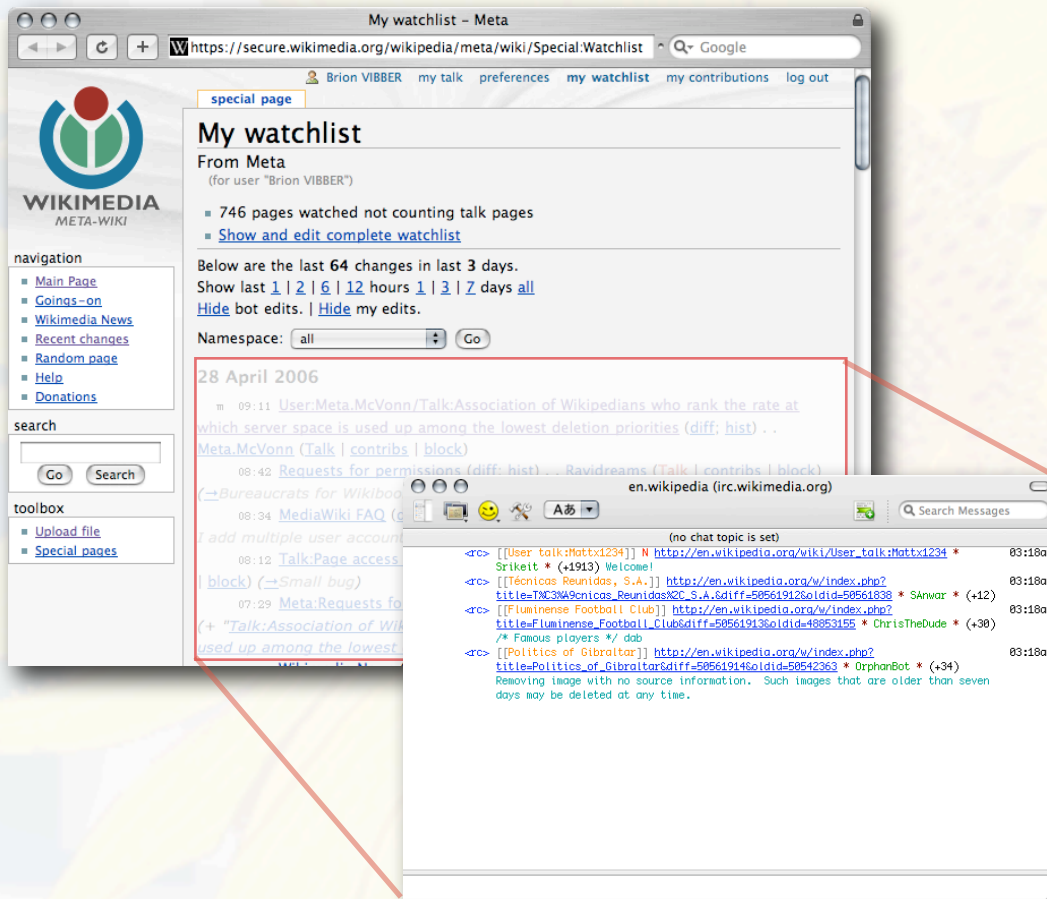
- Cutting out uninteresting things makes cool stuff easier to see

RSS, watchlist: filtering space in recent time

- filter by topic and timeliness, but there's so much more we could do:
- todo: watch keywords
- todo: watch by similarity
- todo: watch by heuristics



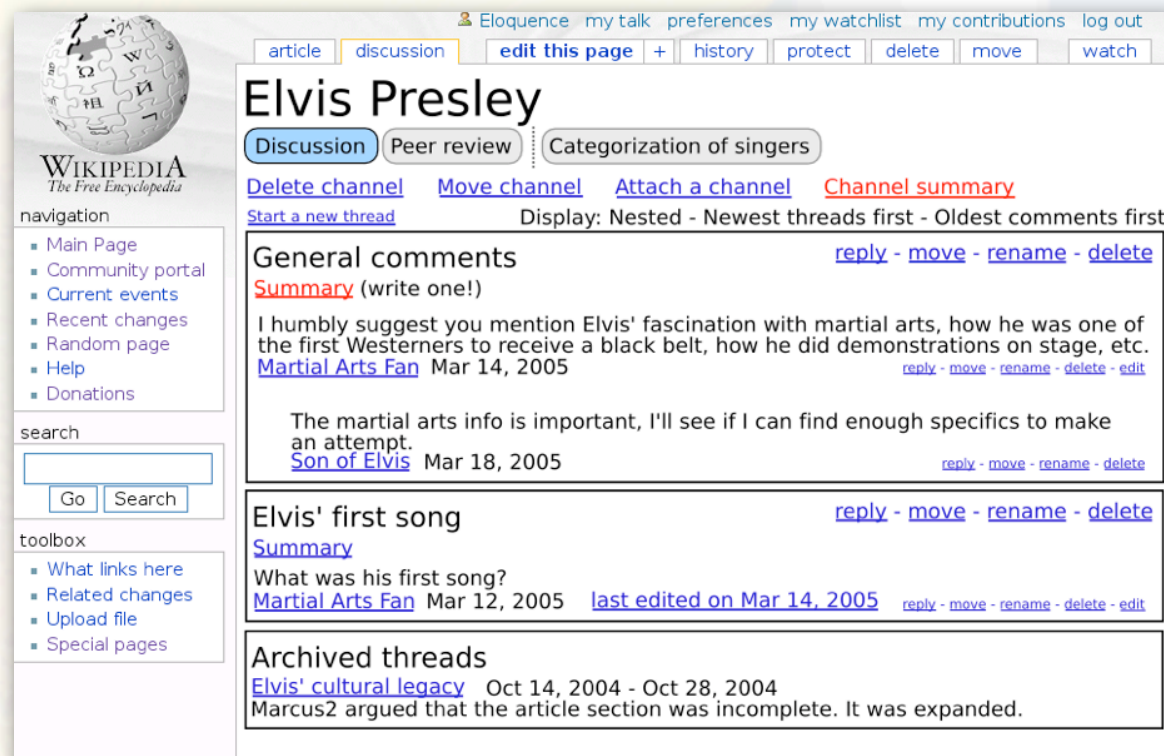
Filtering live time?



- Now: live IRC data streams
- Now: junk edit detection
- Todo: streaming AJAX?
- Todo: desktop notification applets?

Filter discussion space-time

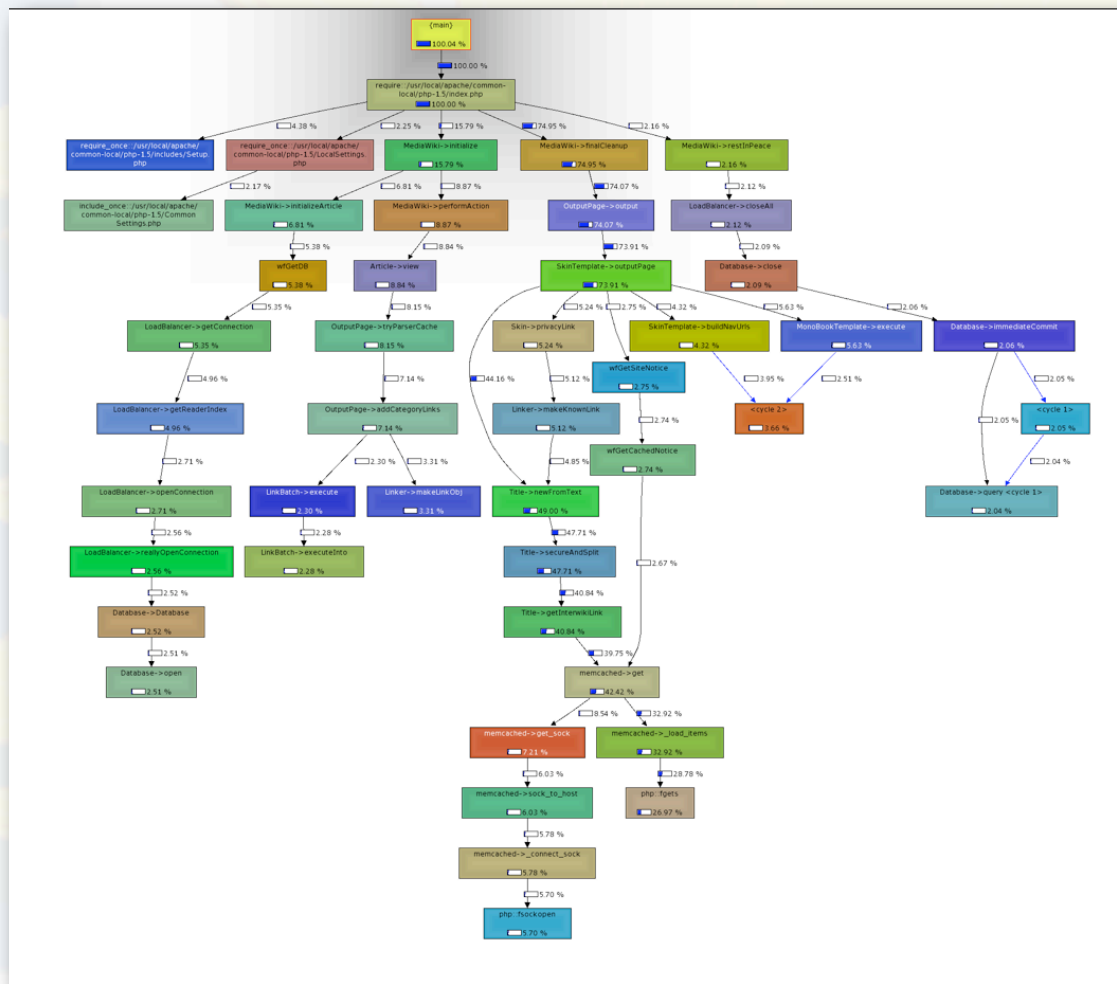
- Active discussions float to top
- Old issues float to bottom
- but... old questions unanswered remain relevant





Interactive time

Where does the time go?



- Profile, profile, profile!
- PHP, filesystem, database, cache...
- As time goes by, request overhead approaches 0
- Well, let's see how close we can get...



Assume software is instantaneous.
What's left?

186,000 miles per second:
It's not just a good idea, it's the law!

War on Latency: Network

Amsterdam



15ms



Frankfurt

Caching proxies are awesome!

- Very fast for static, public page views
- Geographically distributed caches reduce network latency

Tampa



War on Latency: Network

Amsterdam



130ms



Frankfurt

115ms

Tampa



BUT proxies are dumb...

- Login cookie? Page changed since last view?
- Requests now cross an ocean...
- 100-400ms of dead time, depending on continent. Yuck!

Vertical partitioning?

Amsterdam



15ms

130ms



Frankfurt

115ms

Tampa



- Low-latency access to local read-only database slaves
- Transparently pass write operations up the chain
- Provides read-only view if core servers unreachable

More work to keep data synchronized and caches fresh

Who's afraid of a little work?

Smart caches can fail gracefully

On our budget, two nines sound pretty good.
But we would love to do better...

Aggressive AJAX?

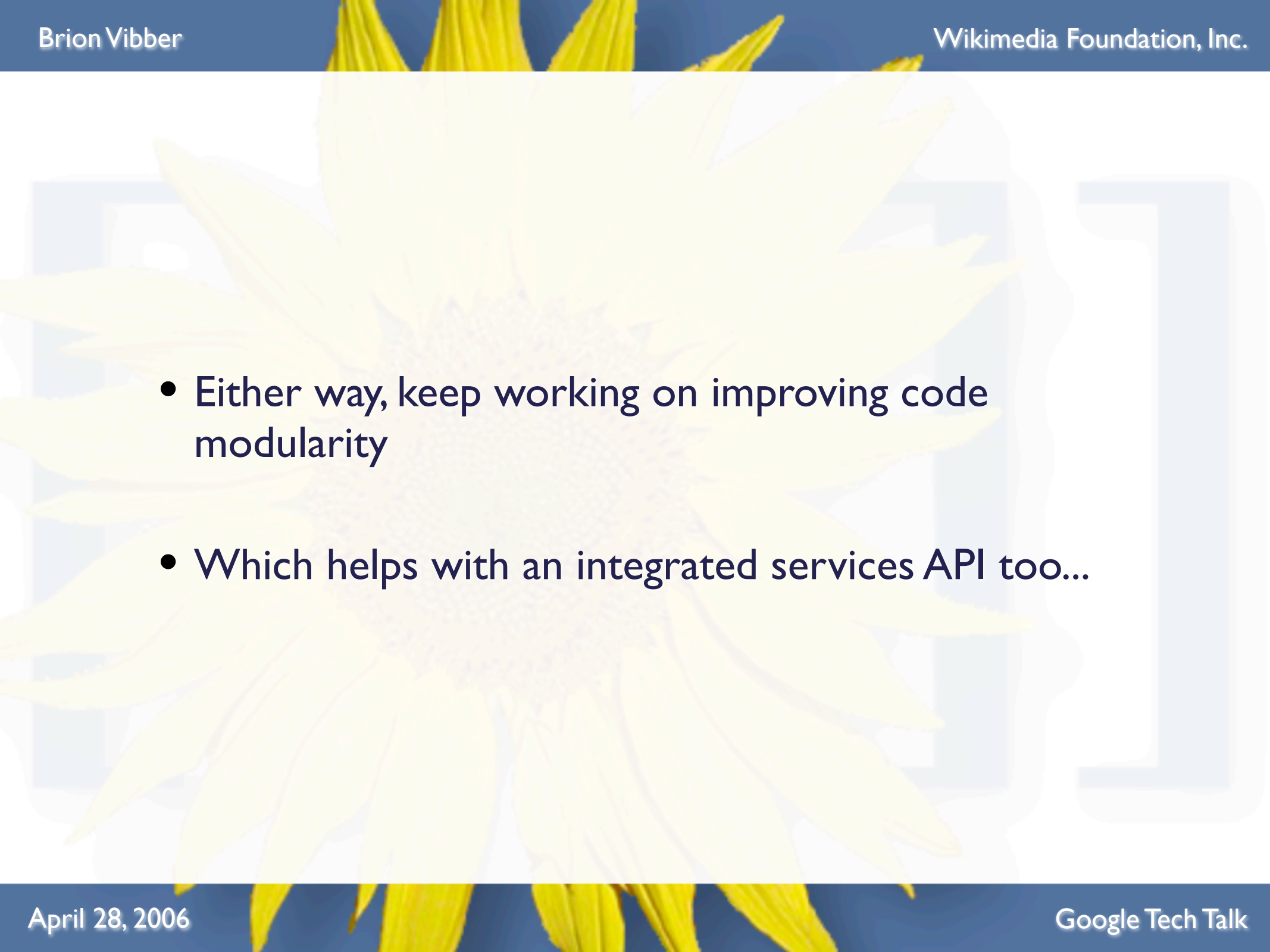
- Run most user login / customization through JavaScript
- Keep most real data in cache-friendly static responses

BUT...

Harms accessibility, compatibility.

Likely to use in moderation.



- 
- Either way, keep working on improving code modularity
 - Which helps with an integrated services API too...

Integration and Outreach

Beta bubble buzzword bingo!

Identity is theft

- Signing up for accounts everywhere sucks!
- Wikipedia and its sister projects to merge user accounts soon
- OpenID, Yadis etc will allow consistent identity on wikis and blogs run by different operators

Web services API

- Screen-scraping sucks
- Many fun uses for querying data from Wikipedia, Wiktionary, or Wikimedia Commons
- Current work targeting maintenance tools, but content is coming...

Put the Media in MediaWiki

- Wikimedia Commons stores photos, sound, and video clips for Wikipedia & friends
- GNU FDL, Creative Commons, Public Domain
- Planning a handy service for remote use in off-site wikis and blogs etc
- Steal good APIs - Google, Flickr, etc

